图书情报工作
第62卷 第18期 2018年9月

基于学科规范引文影响力与同行评议相关性的科学评价实证研究*

宋丽萍[1] 王建芳[2]
[1] 天津师范大学管理学院 天津 300387 [2] 中国科学院科技战略咨询研究院 北京 100190

摘要：［目的/意义］分析学科规范引文影响力在科学评价中的可行性及其与同行评议的相关性，为负责任计量及以其为支撑的同行评议提供借鉴。［方法/过程］选取 F1000 以及 InCites 平台，将 29 850 篇细胞生物学文献、30 326 篇生物技术文献的 CNCI（学科规范化引文影响力）与被引频次进行相关分析，对其中 956 篇细胞生物学论文的 CNCI 与 F1000 分值进行斯皮尔曼相关系数检验。［结果/结论］研究结果表明，从统计学视角看 CNCI 与被引频次呈高度正相关，与 F1000 分值呈显著正相关，同时亦存在二者相悖的情形。因此，CNCI 在一定程度上能够反映同行评议结果、能代偿实施学术影响力归誉的功能，并适用于跨学科比较；但同行评议或 CNCI 单独作为科学评价标准都会有失偏颇，以 CNCI 为代表的新一代负责任计量指标为支撑的同行评议将成为未来科学评价的主流。
关键词： 同行评议 负责任计量 F1000 InCites
分类号： G250
DOI：10.13266/j.issn.0252-3116.2018.18.013

2014 年 12 月 18 日，备受瞩目的英国科研卓越评价体系 REF 评审结果落下帷幕。2015 年 7 月，由英国科学计量学家、大学管理人员、公共政策专家共同撰写的题为《计量潮：研究评价与管理指标作用的独立审查报告》（The metric tide：report of the independent review of the role of metrics in research assessment and management）问世。报告中提出负责任计量（Responsible metrics）与知情同行评议（informed peer review）的计量原则，并特别指出，科学评价倍受学科因素困扰，建议引入学科规范化引文影响力指标[1]。因此，本文选取 InCites 平台以及 F1000 数据库，探讨学科规范化引文影响力（Category Normalized Citation Impact，CNCI）在科学评价中的可行性及其与同行评议的相关性，以为负责任计量及知情同行评议理念的实施提供借鉴。

1 相关背景

引用影响力是科学计量、乃至科学评价中的主角之一，然而不同学科、出版年、文献类型严重制约了引用影响力的测定。E. Garfield 曾指出：不同学科原始被引频次不能直接进行比较[2]。《莱顿宣言》则指出：不同学科引用差异显著，数学期刊最高影响因子为 3，细胞生物学却高达 30[3]。鉴于引用密度（citation density）的差异，对于不同学科、不同出版年以及不同文献类型引用指标进行规范化成为科学计量的重要课题。

1986 年，T. Braun 和 A. Schubert 在 *Scientometrics* 上发表《不同学科科研成果及其引用影响力之评价》一文，其中提出建立跨学科比较的相对指标，自此文献计量领域开始规范化指标的探索与实践[4]。2010 年，H. F. Moed 提出单篇文章源标准影响指数（Source Normalized Impact per Paper，SNIP），其中从篇均引文数的角度，借鉴引用潜力和标准化的思想修正了不同学科期刊引用行为的差异，从而实现了不同学科领域间期刊的直接比较。近年来，伴随着英国高等教育从基于同行评议的 RAE 到以计量为主体的 REF 的变革，计量潮席卷全球，科学计量从情报学分支进化为科技政策制定乃至科研管理的重要工具，多学科评价成为当务之急，因此，汤森路透科技集团（现科睿唯安）在其科研评价平台 InCites 中推出 CNCI，以应不同学科学术影响

* 本文系国家社会科学基金一般项目“负责任计量视角下科学评价方法与指标优化研究”（项目编号：18BTQ074）研究成果之一。
作者简介：宋丽萍（ORCID：0000-0003-3056-3720），教授，博士；王建芳（ORCID：0000-0002-9950-7818），副研究员，博士，通讯作者，E-mail：wangjianfang@casisd.cn。
收稿日期：2017-11-11 修回日期：2018-06-26 本文起止页码：122-128 本文责任编辑：徐健

力比较之需。

作为本文比较对象的 F1000(即 Facaulty of 1000)，为生物医学领域后同行评议(post-publication peer review)平台。2002 年由 V. Tracz 基于由千名科学家构成虚拟团队的理念而创立。2015 年，在痛诉传统科学出版"五宗罪"之余，V. Tracz 将 F1000 整合为三项独特的服务：即 F1000Prime、F1000Research 以及 F1000Posters，从而集开放获取期刊、开放数据、开放同行评议、开放科学于一体，并保证了研究成果的及时传播[5]。其中，F1000Prime 是目前世界上唯一依据8 000多名生物学和医学领域顶尖科学家同行评议结果对最重要的科研论文提供在线推荐服务的平台，F1000 因子(F1000 Score，FS)为专家评议分值。因此，F1000不仅是检索工具，同时也是重要的科学评价工具。

本文以 FS 为准绳，通过 FS 与对应论文 CNCI 的相关分析，探索 CNCI 为代表的新一代计量指标在科学评价中的适用性。

2 数据的采集与统计

作为新一代研究分析平台，InCites 数据库基于汤森路透 Web of Science 核心合集七大索引数据库而建立。其中一篇文献的 CNCI 是这样定义的：一篇文献实际被引次数除以同文献类型、同出版年、同学科领域文献的期望被引次数[6]。其计算公式为：

$$CNCI = \frac{x}{e_{ftd}}$$

当一篇文献被划归至多个学科领域时，则使用每个学科领域实际被引次数与期望被引次数比值的平均值，其计算公式为：

$$CNCI = \frac{\sum \frac{c}{e_{f(n)td}}}{n} = \frac{\frac{c}{e_{f(1)td}} + \frac{c}{e_{f(2)td}} + \cdots + \frac{c}{e_{f(n)td}}}{n}$$

这里，e = 期望引用率或基线值，c = 总被引次数，f = 学科领域，t = 出版年，d = 文献类型，n = 论文被划归的学科领域数。

如果 CNCI 的值等于 1，表明该篇论文的被引表现与全球平均水平相当，CNCI 大于 1，则表明该篇论文的被引表现高于全球平均水平；小于 1，则低于全球平均水平。从这一意义上说，CNCI 排除了出版年、学科领域与文献类型的影响，是一个具有实用价值的影响力指标。

InCites 平台检索表明，基于 Web of Science 252 个不同学科分类的论文、综述的被引频次差异显著，从材料科学的754 184 次到诗歌艺术的51 次。图 1、图 3 为2014 年不同学科论文、综述被引频次与学科规范化引文影响力差异性对比图，图 2、图 4 为不同学科论文、综述篇均被引及学科规范化引文影响力差异性对比图，其中横轴代表不同学科，纵轴代表被引频次或学科规范化引文影响力数值。因此表明在科学评价实践中引入 CNCI 指标的必要性。对此，L. Bornmann、L. Leydesdorff 等学者进行了深入分析和验证[7-8]。

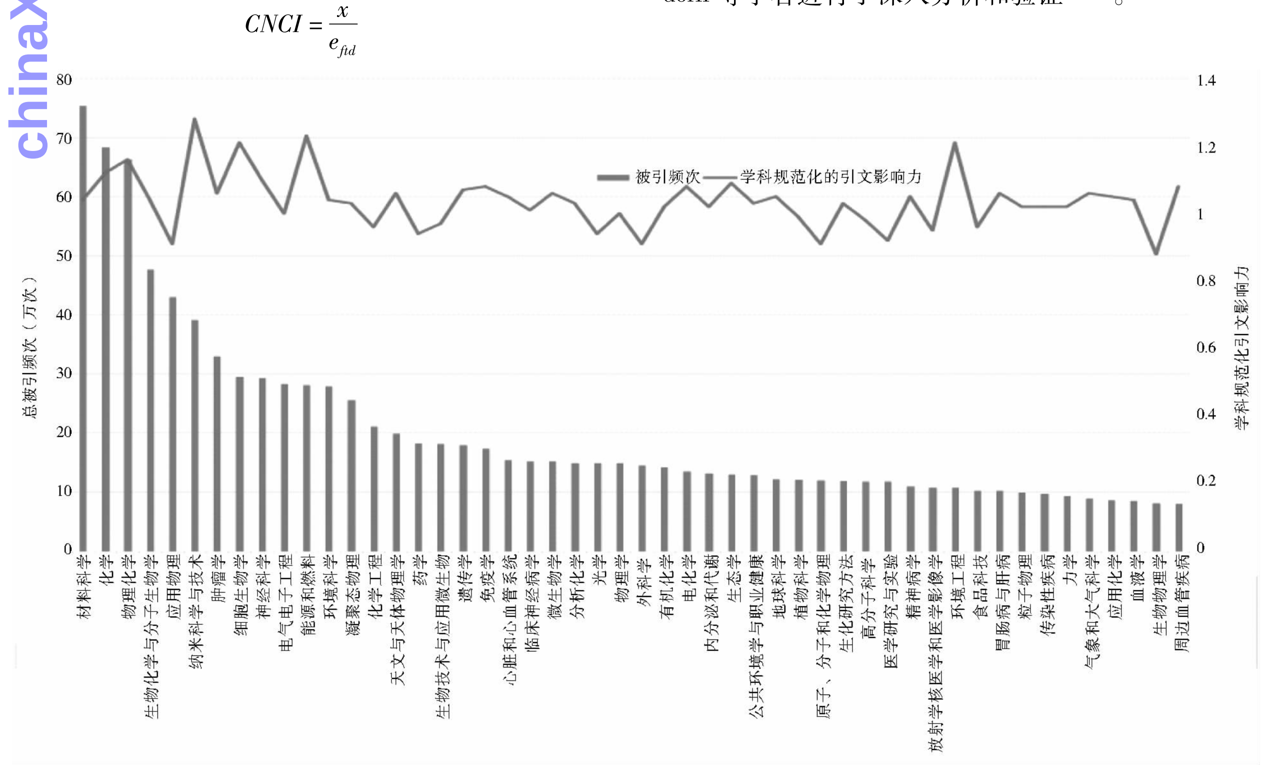


图 1 不同学科论文被引频次及学科规范化引文影响力的差异性对比

图书情报工作
第 62 卷 第 18 期　2018 年 9 月



图 2　不同学科论文篇均被引频次及学科规范化引文影响力的差异性对比

图 3　不同学科综述被引频次及学科规范化引文影响力的差异性对比

本文确定 2014 为统计年，以保证足够的统计被引区间。数据采集步骤如下：①确定统计对象：在 InCites 平台提供的 9 种学科分类体系中，本文选取常用的 Web of Science 分类，将 Web of Science 的 252 个分类与 F1000Prime 高级检索界面中 40 个研究方向进行对比，从而确定二者共同的细胞生物学（Cell biology）、生物技术（biotechnology）为统计对象。②在 InCites 平台中检索。以细胞生物学为例，在筛选项中拟定检索式为 field = cell biology and article publication date = 2014。检索结果为细胞生物学论文 26 137 篇，综述 3 713 篇；生物技术论文 27 660 篇，综述 2 666 篇。继而，检索获得每篇论文的 DOI、标题、作者、被引频次、CNCI、学科领域百分位、期刊规范化引文影响力、期刊影响因子等指标[9]。③在 F1000Prime 中进行高级检索，仍以细胞生物学为例，检索式为 article publication date = 2014 and facaulties = cell biology，检索获得细胞生物学文献

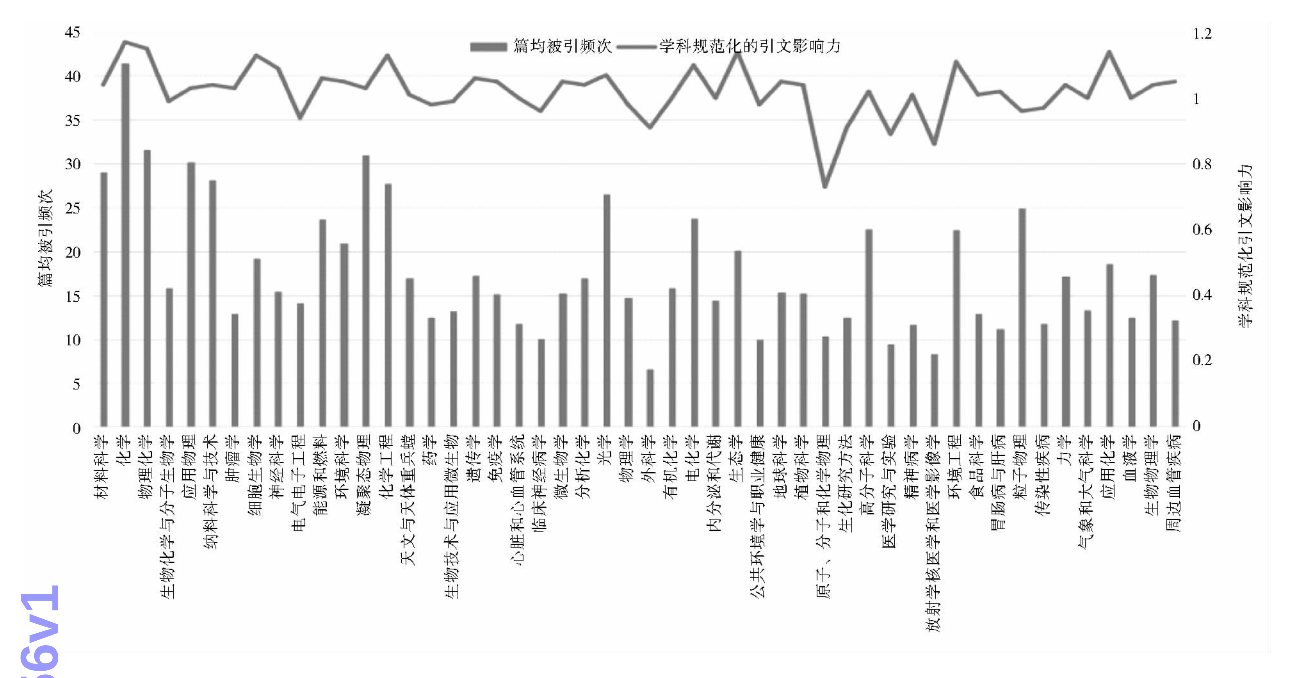


图 4　不同学科综述篇均被引频次及学科规范化引文影响力的差异性对比

3 334篇，生物技术文献 2 225 篇，同时抓取每篇文献的标题、DOI 以及 FS。检索时间为 2017 年 6 月 5 日[10]。④利用 if(countif(,),“S”,“F”)组合函数与 VLOOKUP 函数筛选具有共同 DOI 的文献及其 CNCI、FS 数值。因此获得 956 篇细胞生物学论文，20 篇细胞生物综述，91 篇生物技术论文(生物技术综述数量过少，不具统计意义，略去)用于进一步分析。

3　数据的处理

依据上述统计结果，借助 SPSS22.0 进行统计分析。首先，采用 K-S 检验进行数据正态性分析，以便对数据的分布状态进行探测，并选用适宜的统计学方法。分析表明(过程从略)，显著性水平小于 0.05，因此上述数据集均不服从正态分布，故选用非参数的斯皮尔曼检验为宜。

3.1　单篇论文被引频次与 CNCI 相关分析

26 137 篇细胞生物学论文 CNCI 与被引频次的斯皮尔曼相关分析结果见表 1，27 660 篇生物技术论文分析结果从略。其中 DCF 代表被引频次。

表 1　细胞生物论文 CNCI 与被引频次相关性

			DCF	CNCI
Spearman 的 rho	DCF	相关系数	1.000	.989 **
		显著性(双尾)	.	.000
		N	26 137	26 137
	CNCI	相关系数	.989 **	1.000
		显著性(双尾)	.000	.
		N	26 137	26 137

注：**. 在 0.01 显著性水平上(双尾)

3 713 篇细胞生物学综述 CNCI 与被引频次的斯皮尔曼相关分析结果(2 666 篇生物技术综述分析结果从略)，如表 2 所示：

表 2　细胞生物综述 CNCI 与被引频次相关性

			DCF	CNCI
Spearman 的 rho	DCF	相关系数	1.000	.990 **
		显著性(双尾)	.	.000
		N	3713	3713
	CNCI	相关系数	.990 **	1.000
		显著性(双尾)	.000	.
		N	3 713	3 713

注：**. 在 0.01 显著性水平上(双尾)

3.2　单篇论文 CNCI 与 FS 相关分析

以细胞生物学论文为例，956 篇细胞生物论文 DCF、CNCI 与 FS 斯皮尔曼相关分析结果如表 3 和表 4 所示：

表 3　细胞生物学论文 CNCI 与 FS 相关性

			DCF	CNCI	FS
Spearman 的 rho	DCF	相关系数	1.000	.990 **	.364 **
		显著性(双尾)	.	.000	.000
		N	956	956	956
	CNCI	相关系数	.990 **	1.000	.362 **
		显著性(双尾)	.000	.	.000
		N	956	956	956
	FS	相关系数	.364 **	.362 **	1.000
		显著性(双尾)	.000	.000	.
		N	956	956	956

注：**. 在 0.01 显著性水平上(双尾)

表 4 97 篇细胞生物学论文 CNCI 与 FS 相关性

			CNCI	FS
Spearman 的 rho	CNCI	相关系数	1.000	.163
		显著性(双尾)	.	.110
		N	97	97
	FS	相关系数	.163	1.000
		显著性(双尾)	.110	.
		N	97	97

注:**. 在 0.01 显著性水平上(双尾)

4 结果与讨论

4.1 统计检验结果

依据相关系数|r|≥0.8 为高度相关,0.5≤|r|<0.8 为中度相关,0.3≤|r|<0.5 为低度相关,|r|<0.3 为不相关的界定标准,上述相关性分析结果如下:

(1)无论是细胞生物学亦或是生物技术,无论是论文亦或是综述,CNCI 与被引频次均呈高度正相关,相关系数分别为 0.989、0.990、0.993、0.992。

(2)以 FS 为准绳,在 956 篇细胞生物学、91 篇生物技术论文中,FS 与 CNCI 均呈显著正相关,相关度分别为 0.362、0.443,即 FS 与 CNCI 总体变动方向一致。然而低度相关的结果证明 FS 与 CNCI 两个变量在变动方向总体一致的趋向下,亦存在着不同之处,即存在同行评议与引文分析相左的情形。具体地说,有一些论文专家打分很高然而被引频次、CNCI 较低,反之亦然。

(3)鉴于 FS 与 CNCI 的低度相关,这里以细胞生物学论文为例,进一步选取细胞生物学中 FS 为 5 以上的 97 篇论文进行同行评议与 CNCI 相关性分析(见表 4)。分析表明,二者相悖的趋向愈加显化,相关度由原有的 0.362 下降为 0.163,即从低度相关进入不相关。进一步观察,证实了这一结论的可信性。956 篇论文中,FS 居于第二位的论文,CNCI 为 611 位;反之,CNCI 排名第一的论文,FS 居于 420 位,CNCI 排名第二的论文,FS 仅居 666 位。

4.2 结果分析

(1)基于引用的 CNCI 与基于质量考量的 FS 正相关即说明二者总体目标的一致性,并很好地诠释了国际最新主流理念——知情同行评议。事实上,在科学评价实践中,一方面,指标是量化分析的基本元素;另一方面,指标在简化认识的同时伴随着信息的流失,并且没有一个单一指标可以涉及所有的关键变量。因此,基于指标构建学术影响力框架时,一套指标完胜于单个指标,以取得更好的信度与效度。同时,鉴于与同行评议结果的契合度是衡量指标好坏的重要砝码,通过 CNCI 等指标与开放同行评议结果 F1000 因子的相关性,探索同行评议与 CNCI 结合的方式,将使科学评价结果更逼近评价对象本来面目。

(2)在 FS 与 CNCI 总体一致的趋向下,亦不排除二者相悖的情形,因此进一步说明同行评议或引文分析,单独作为科学评价标准都会有失偏颇。无疑,无论是同行评议还是引用都是令人信赖的,同行评议作为信息"守门人"保证进入交流系统信息的质量,特别是 F1000 通过后同行评议对信息进行了再净化,保证了信息的高度遴选和过滤;CNCI 等基于引用的文献计量指标则描述了知识在科学共同体中的扩散过程。按照 J. Wilsdon 的观点:"目前广为采用的文献计量指标即是引用统计"[1];默顿则指出:"引文是科学系统中用来偿还研究债务的货币"。因此,一般认为:引用是学术影响力的可量度指标。那么,低度相关的结果则标志着科学评价中同行评议和引文分析作为质性评价和影响力揭示维度的不同。因此,二者有机结合的方式,即如何以计量支撑同行评议是未来科学评价的核心议题。

(3)CNCI 立足于消除引用影响力的学科偏态分布,故在跨学科比较方面较被引频次等传统引文分析指标具有一定的先进性。然而,目前 CNCI 困囿于学科分类的敏感性,因此,如何规避其局限性,并选取与被引频次、乃至同行评议结果契合度较高的算法是值得研究的问题。本文中,F1000 与 InCites 相同主题检索的匹配度最高的细胞生物学论文仅为 3.65%。究其原因,一方面二者遴选原则不同所致:即 Web of Science 以期刊的高度遴选著称,而 F1000 以"好论文不问出处"为信条;另一方面说明学科分类对于 CNCI 影响甚大。因此,InCites 平台提供 Web of Science、ESI、法城手册(Frascati)、英国的 RAE、澳大利亚的 ERA 等 9 种学科分类体系。2014 年,英国 REF 的实施将 CNCI 从理论探索推向实际应用,分学科评价提上议事日程。在我国,武夷山等针对科研评价中的问题也早已提出同类比较,即分类管理、分类评价的计量原则。2015 年,中共中央办公厅国务院办公厅出台的《深化科技体制改革实施方案》,2016 年,教育部社会科学委员会发布的《高等学校哲学社会科学研究评价指南》,皆提出

建立分类评价制度。由此可见,分类评价已成为国内外科学共同体的共识。然而,分类体系在一定程度上干扰着 CNCI 计算结果。为此,计量学家在提醒人们选择合适的分类体系的同时,积极探索 CNCI 的变体,以统一样本的统计分布性、更好地消除引用影响力的学科偏态分布造成的不可比性。L. Waltman、L. Bornmann、L. Leydesdorff、H. F. Moed、M. Thelwall 等学者开发的诸多学科规范化算法,S. Haustein 等提出的相对引用比(Relative Citation Ratio)即是在此方面做出的尝试[11]。

(4)同行评议与 CNCI 有机结合使《莱顿宣言》和负责任计量的思想具体化。莱顿宣言的总体原则即是保持以 FS 为代表的质性评价第一位。与此同时,学科规范化已经作为十大原则之一列入《莱顿宣言》成为科学评价的纲领之一。从根本上说,科学评价的主要目的就是鉴别、促进和扶植卓越,科学计量则是以上述目的为核心而建立的一系列指标与方法。CNCI 则代表着以负责任计量为主导思想而建立的科学评价指标,从而保证基于计量的同行评议的付诸实施。从这一意义上说,以 CNCI 为代表的新一代负责任计量指标为支撑的同行评议将成为未来科学评价之进路。

5 结论

通过上述分析,得出结论:

(1)CNCI 在跨学科比较方面具有一定的先进性,可以代偿执行引用统计的信息过滤与学术影响力归誉的功能。但是,鉴于其对于学科分类的敏感性,如何优化 CNCI 目前归一化算法(Mean Normalised Citation Score),以便在不改变数据性质和相对关系的前提下,更好地适应不同学科引文的偏态分布是探索中的问题。

(2)在科学评价实践中,鉴于同行评议的主观性和"马太效应",基于引用的指标可以用来辅助同行评议,以实现更深入的公众参与、体现科研评价的民主化,并提升同行评议质量、降低其偏见与风险[12]。然而,如何将二者有机结合是探索中的问题。CNCI 代表了一种尝试和趋向,其迎合了负责任计量分学科评价之需,一定程度上能够反映同行评议结果,从而为同行评议提供有力支撑。

(3)CNCI 等文献计量指标与同行评议有机结合的方式将是未来科学评价的主流,从而保证基于计量的科学评价的付诸实施。

参考文献:

[1] WILSDON J. The metric tide: report of the independent review of the role of metrics in research assessment and management[M]. London :SAGE Publishing,2016:137-158.
[2] GARFIELD E. Citation indexing - its theory and application in science, technology, and humanities[M]. New York: John Wiley & Sons,1979.
[3] HICKS D, WOUTERS P, WALTMAN L. Bibliometrics:the Leiden Manifesto for Research metrics[J]. Nature. 2015,520 (7548): 429-431.
[4] HAUNSCHILD R, BORNMANN L. Normalization of Mendeley reader counts for impact assessment[J]. Journal of informetrics, 2016, 10(1):62-73.
[5] TRACZ V, LAWRENCE R. Towards an open science publishing platform[EB/OL]. [2018-07-20]. https://f1000researchdata. s3. amazonaws. com/manuscripts/8575/0c0c20d6 - 8920 - 4d1c - b564 - e81901b268f4_7968_-_rebecca_lawrence. pdf? doi = 10. 12688/f1000research. 7968. 1.
[6] InCites[EB/OL]. [2018-04-17]. https://clarivate. com. cn/products/incites/.
[7] LEYDESDORFF L, RADICCHI F, BORNMANN L, et al. Field - normalized impact factors (IFs): A comparison of rescaling and fractionally counted IFs[J]. Journal of the American Society for Information Science & Technology, 2013, 64(11):2299-2309.
[8] BORNMANN L, HAUNSCHILD R. Measuring field-normalized impact of papers on specific societal groups: an altmetrics study based on Mendeley data[J]. Research evaluation, 2016, 26(3): 10-35.
[9] InCites[EB/OL]. [2017-06-03]. https://incites. thomsonreuters. com/.
[10] F1000[EB/OL]. [2017-06-05]. http://f1000. com/prime.
[11] BORNMANN L, HAUNSCHILD R. Relative Citation Ratio (RCR): an empirical attempt to study a new field-normalized bibliometric indicator[J]. Journal of the Association for Information Science & Technology, 2017, 68(4):1064-1067.
[12] DERRICK G E, PAVONE V. Democratising research evaluation: achieving greater public engagement with bibliometrics-informed peer review[J]. Science & public policy, 2013, 40(5):563-575.

作者贡献说明:

宋丽萍:确定研究思路,采集数据,撰写论文主体;
王建芳:相关文献收集、统计分析。

An Empirical Study of Scientific Evaluation Based on the Correlation Between CNCI and Peer Review

Song Liping[1] Wang Jianfang[2]
[1] Management School of Tianjin Normal University, Tianjin 300387
[2] Institutes of Science and Development, Chinese Academy of Sciences, Beijing 100190

Abstract: [**Purpose/significance**] In order to make some recommendations for responsible metrics and peer review, the paper analyses the feasibility of Category Normalized Citation Impact (CNCI) and its correlation with peer review. [**Method/process**] By choosing F1000 and InCites, this paper conducted the correlation analysis between CNCI and citation counts of 29,850 cell biology papers and 30,326 biotechnology papers, and performed the test of spearman rank correlation coefficient between the CNCI and F1000 scores of 956 cell biology papers. [**Result/conclusion**] The results show that CNCI is positively correlated with citation counts, and significantly correlated with F1000 rating, but there are contradictory situations for the two indicators at the same time. Therefore, CNCI can reflect the results of peer review, substitute for the indication of academic influence, and also can be used for comparision among disciplines. However, peer review or CNCI will be biased for scientific evaluation purpose alone. Then, tools that link responsible metrics indicators, such as CNCI and informed peer reviews would be valuable additions to the field of research evaluation.

Keywords: peer review responsible metrics F1000 InCites